How do you define the value of something if it's free? Observations on Caltech's Institutional Repository

Donna Wrublewski
Chemistry & Biological Sciences Librarian

George Porter
Interim Head, Research Services
Engineering Librarian

California Institute of Technology

Defining "Value" in Scholarly Communications:
Evolving Ways of Evaluating Impact on Science

249th American Chemical Society Meeting
Denver, Colorado

March 22, 2015
CINF-9

It's a pleasure to be a part of this because one of my favorite conversations to have with people in the field is about the relationship (and decoupling) between price, cost and value. The previous talks have looked at the impact of data and personal or social marketing, and I look forward to the panel discussion. My presentation has a unique aspect on this in that it is, to a degree, a mix of all of these, plus one that hasn't been specifically addressed yet – trying to determine impact from the point of view of the Academic Institution. I want also to thank my co-author on this presentation, George Porter, who was a driving force behind the creation and expansion of our repository.

California Institute of Technology

AAU member

~2200 students, ~300 faculty, 600 researchers

6 Divisions, ~25 Options

- Biology & Biological Engineering
- Chemistry & Chemical Engineering
- Engineering & Applied Sciences
- Geology & Planetary Sciences
- Humanities and Social Sciences
- Physics, Math and Astronomy

150-200 PhD dissertations per calendar year

Sherman Fairchild Library

One of 5 libraries on campus

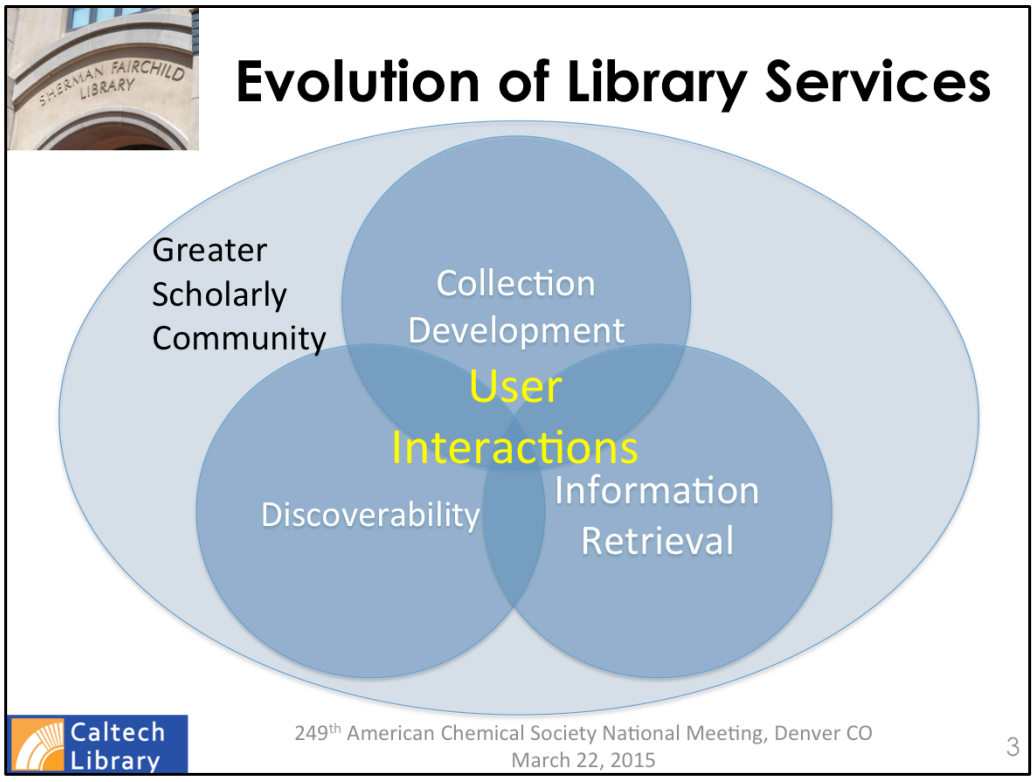
Opened in 1997, 24 hours

Serves all science & engineering disciplines

410,000 monographs (~7% electronic)

2,100 journals (90% electronic)

First a little background about Caltech and the Library. The past few years have seen between 150 and 200 PhD dissertations granted per year by the Institution. We currently have 6 subject librarians plus our Repository Librarian. We are a small operation, and thus strategically choose our initiatives.



My last ACS talk in San Francisco focused on “internal” library initiatives serving primarily Caltech faculty and students. At the very end, I briefly mentioned our Institutional Repository, and its role in serving the Greater Scholarly Community. I’m pleased to be able to talk now more about how this expands to the world outside of Caltech in terms of the world accessing our materials, and how we assist our researchers with communicating with “the world” in terms of information promotion and legal compliance. I’m going to describe our repository and our services, and then towards the end, look at this service and show how it ties in with our institutional mission.

Caltech Collection of Open Digital Archives (Caltech CODA)

- Consists of several different repositories, active since 2001
- **CaltechAUTHORS (47K records)** – scholarly output (articles, book chapters, books, individual conference papers, etc.)
- **CaltechTHESIS (8K records)** – Primarily Ph.D. Dissertations, some M.S. and Undergraduate
- CaltechCampusPubs, CaltechOH, CaltechLabNotes, CaltechES, CaltechCONF
- http://libguides.caltech.edu/CODA

Caltech Library

249th American Chemical Society National Meeting, Denver CO
March 22, 2015

4

The Caltech Institutional Repository is formally known as Caltech CODA – the Caltech Collection of Open Digital Archives. Started in 2001, it is comprised of several different repositories. I will focus on two of them today – CaltechAUTHORS and CaltechTHESIS. Other repositories are CampusPubs which contains our student newspaper and yearbooks, OH which has alumni oral histories, LabNotes which has notebooks and research papers, ES which contains the Engineering & Science magazine, and CONF which houses conference proceedings. More information can be found on our CODA LibGuide.

CaltechAUTHORS

- "On March 16, 2009, the Faculty Board unanimously approved a motion that Caltech Faculty support the creation of the Caltech Collection, and that the library be charged to create and maintain this collection."
- Maintained by Library Staff
  - Weekly literature searches in Web of Science, SciFinder, PubMed & MathSciNet
- Enhanced metadata including grant numbers, PMCIDs, OA Status, and (recently) ORCIDs
- "Historical" projects as requested

http://libguides.caltech.edu/CODA
http://libguides.caltech.edu/CaltechOAPolicyFAQ
http://library.caltech.edu/coda/OA_Policy_6.10.2013.pdf

Caltech Library

249th American Chemical Society National Meeting, Denver CO
March 22, 2015

5

CaltechAUTHORS was initially launched in 2001 and populated two ways. Motivated individuals (a few faculty and post docs) deposited their own work, and library staff harvested papers based on the ability to make the Version of Record freely available on behalf of the author. In 2009, CaltechAUTHORS was designated by the Faculty Board as the official repository for Caltech scholarly output. Today, staff are trained to perform weekly address/affiliation searches of Web of Knowledge, SciFinder, PubMed, and MathSciNet to search for citations and enter the relevant metadata and files into the repository. Librarians review submissions before posting. Historical "backfile" projects to complete scholarly records of faculty are done as requested and as time permits otherwise.

CaltechAUTHORS and Caltech's Open Access Policy

- Caltech Open Access policy adopted June 10, 2013; effective January 1, 2014
  - "The Faculty of the California Institute of Technology is committed to disseminating the fruits of its research and scholarship as widely as possible."
  - "…each Faculty member grants to the California Institute of Technology a nonexclusive, irrevocable, worldwide license to exercise any and all rights under copyright relating to each of his or her scholarly articles…"

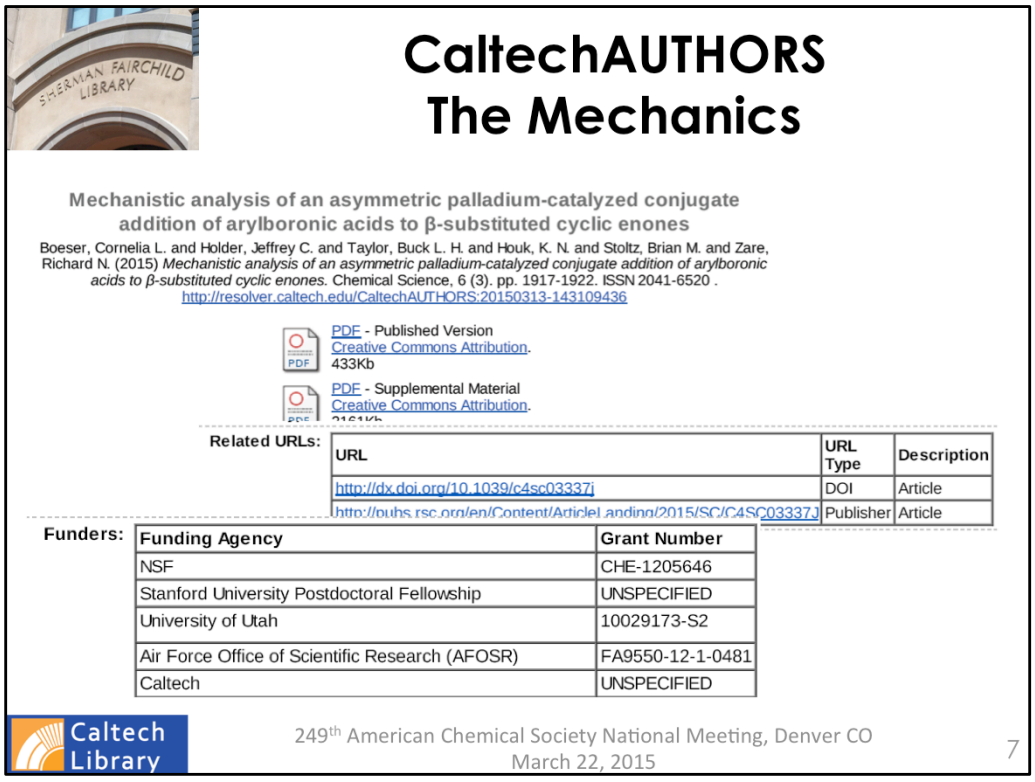
http://libguides.caltech.edu/CODA
http://libguides.caltech.edu/CaltechOAPolicyFAQ
http://library.caltech.edu/coda/OA_Policy_6.10.2013.pdf

Caltech Library

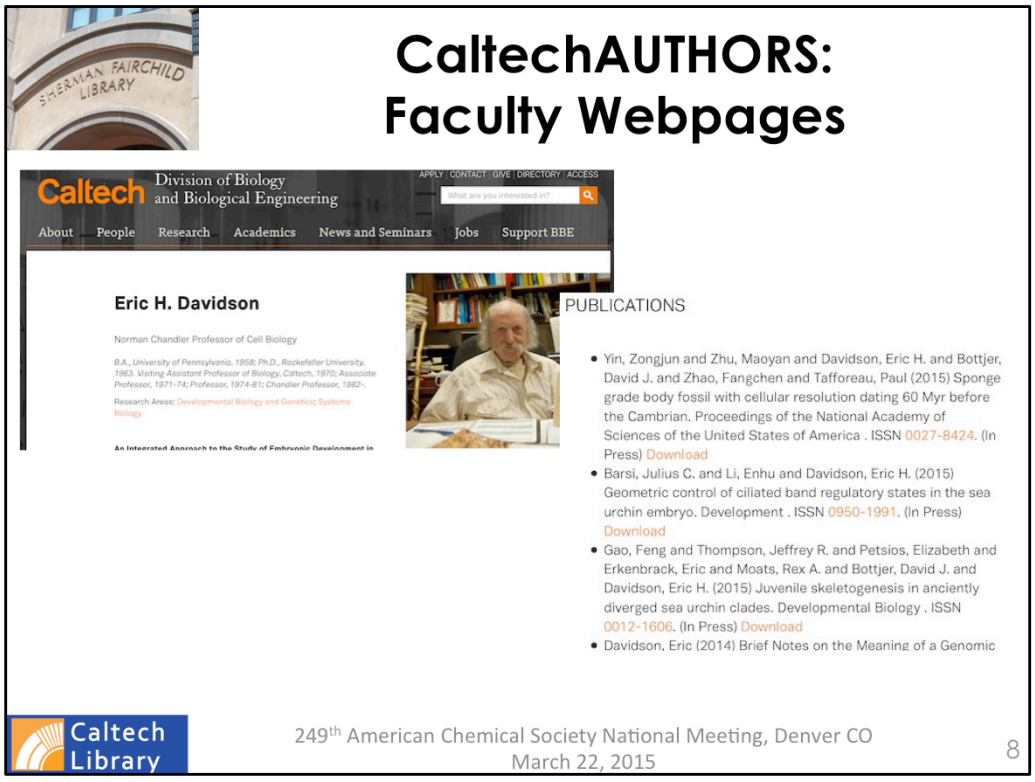
249th American Chemical Society National Meeting, Denver CO
March 22, 2015

6

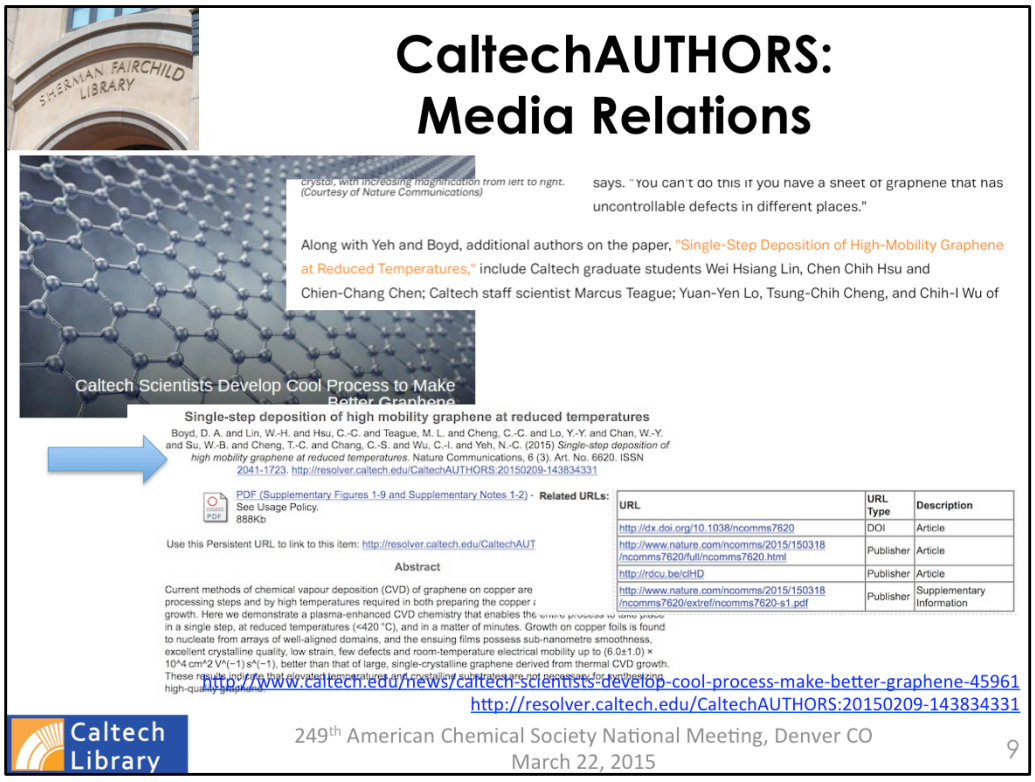
CaltechAUTHORS allows faculty to comply with the University's open access policy, and allows for manuscripts and/or final PDF copies of papers to be made publicly available in compliance with copyright. Waivers are granted when needed for articles appearing in non-cooperating journals (such as Nature).



The deposit process involves filling out a form and attaching articles and supplementary information. These are made publicly available when possible (for example when an article has a CC license or an author manuscript version is allowed to be posted). Metadata is added, including author names and identifiers, links to articles & supplementary information, and funding information. Administrative staff often request training on how to deposit for their faculty.



Caltech introduced a new design and layout for the campus websites, which included an option to link directly to CaltechAUTHORS to display an up-to-date list of a faculty member's publications. There are also advanced options for faculty to create and display custom lists. From my time at the University of Florida, I know this can be an important "marketing" tool (insomuch as Caltech needs marketing) – no one like to see a website that is outdated and a publication list that stopped in 2007. CaltechAUTHORS makes this extremely easy.



Recently the Library has partnered with the Caltech Media Relations department to provide ongoing support to enhance Caltech Press Releases with direct links to papers, including full text where possible. A recent paper on graphene is an example of this. The work is highlighted on the Caltech homepage, and the link in the press release to the scientific article takes a user to the citation in CaltechAUTHORS. From here the user can freely view the abstract, grant and author information, and other metadata. If the article is freely available in CaltechAUTHORS, a direct link to it is provided. Links to publisher pages and other repositories are also provided, and a user who has a subscription to the journal can access it directly through them.

CaltechAUTHORS: Global Reach

- Back-end search to retrieve articles with keywords of interest
- Use article Eprint ID numbers to retrieve access IP addresses
- Map global access information using OpenHeatMap.com

Biophotonics
Nanofluidics
Nanophotonics
Natural Products
Organometallics

Caltech Library

249th American Chemical Society National Meeting, Denver CO
March 22, 2015

10

As of now AUTHORS has just over 47,000 records, and almost 6 million downloads since July 2008. There has also been a significant increase in referrals from the Caltech homepage due to the Media Relations effort. This is an access ratio of about 127:1 (6M downloads / 47K articles). So, in terms of impact, we can see that people are looking at our content (but like Tony said, views don't necessarily translate to impact). Eprints currently allows heatmap generation for individual articles or authors, but not by a topic. We're curious about the global reach of papers on the specific topics listed here.

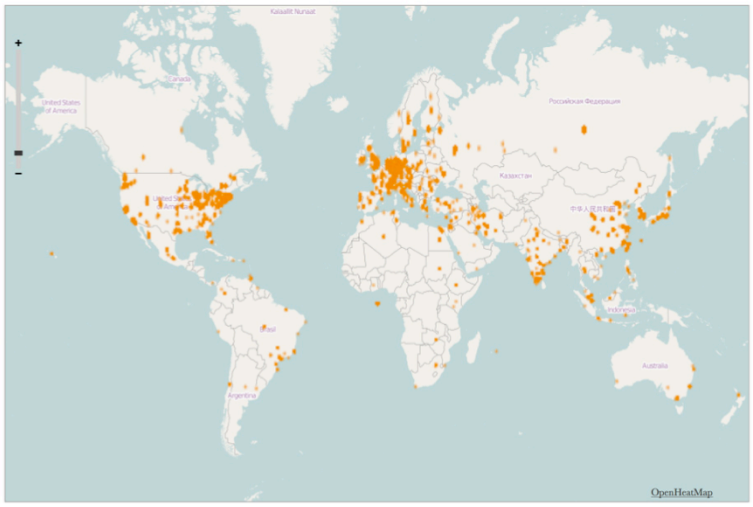
4 papers, ~2700 accesses

To get an idea of the global reach of CaltechAUTHORS we used the following procedure to analyze access data. The keywords indicated were searched against records in Eprints, matching on Title, Abstract, or Keyword, and the relevant articles were identified. The access records for those articles were then downloaded, and the IP's were uploaded to OpenHeatMap.com to generate the maps.

CaltechAUTHORS: Global Reach and "Impact"

Nanofluidics

OPENHEATMAP



Caltech Library

249th American Chemical Society National Meeting, Denver CO
March 22, 2015

12

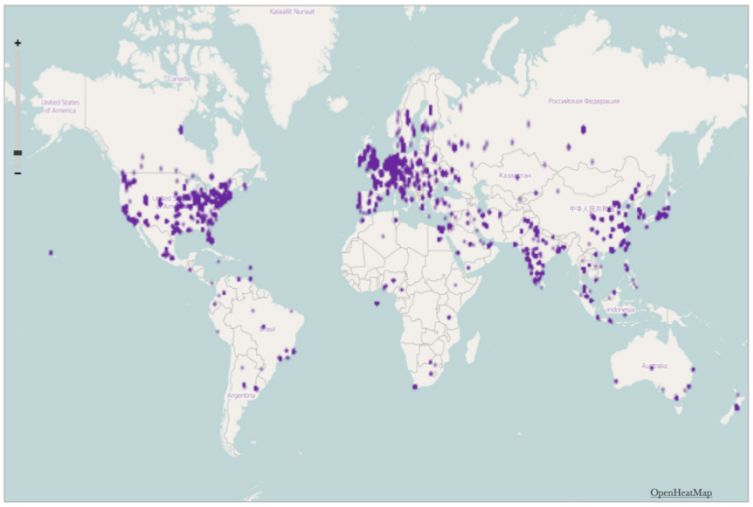
8 papers, ~19,000 accesses

34 papers, ~39,700 accesses

CaltechAUTHORS: Global Reach and "Impact"

Natural Products

OPENHEATMAP

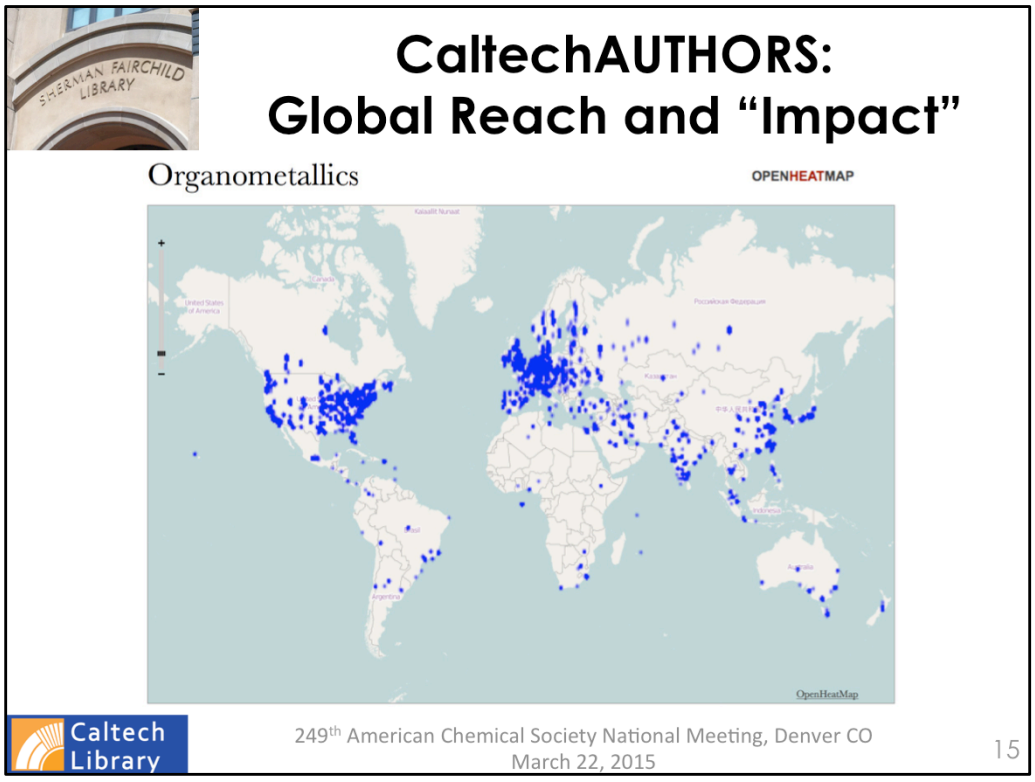


Caltech Library

249th American Chemical Society National Meeting, Denver CO
March 22, 2015

14

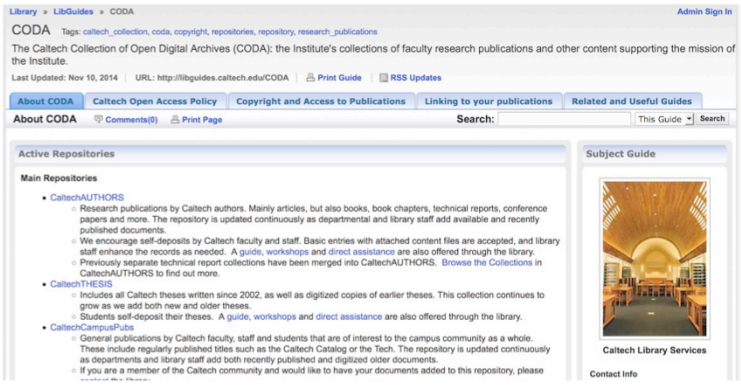
50 papers, ~41,300 accesses



46 papers, ~37,700 accesses

CaltechAUTHORS: Outreach

- Classes
  - For the Community
  - For Staff



http://libguides.caltech.edu/CODA

Caltech Library

249th American Chemical Society National Meeting, Denver CO
March 22, 2015

16

George and Kathy teach classes on CaltechAUTHORS. We have two flavors – one for the general community to learn about what's in it and how they can contribute, and one specifically for faculty support staff. The staff one focuses on depositing, and is usually of interest to those faculty who want to have their complete publication record in CaltechAUTHORS to use on their website, and also going forward for things like the new NIH Biosketch Bibliography section. This was a very recent development in which I worked with the NIH to have institutional repository bibliographies allowed as acceptable links in the Biosketch, and I would be happy to discuss that in the Q&A or later on during the conference.

CaltechTHESIS

- Voluntary in 2001, mandatory for all candidates as of July 2002
- Ability to restrict when needed

Synthetic applications and methodological developments of donor–acceptor cyclopropanes

Citation

Goldberg, Alexander F. G. (2013) *Synthetic applications and methodological developments of donor–acceptor cyclopropanes.* Dissertation (Ph.D.), California Institute of Technology.
http://resolver.caltech.edu/CaltechTHESIS:02202013-172928174

Abstract

Donor–acceptor cyclopropanes are a versatile class of synthetic intermediates ring-opening reactions and formal cycloadditions, and employed in numerous have developed new Lewis acid mediated cycloadditions for the synthesis of fi

Thesis Files

PDF (1_title_intro)
Restricted to Caltech community only until 29 May 2015.
See Usage Policy.
308Kb

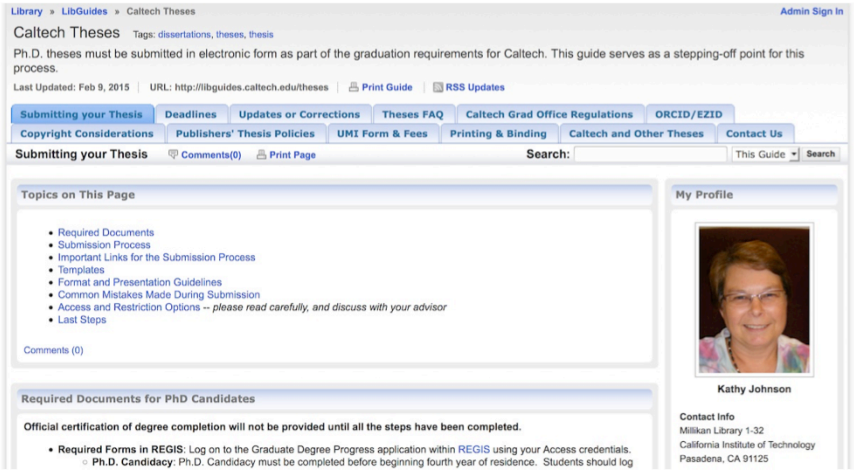
PDF (14_about_the_author)
Restricted to Caltech community only until 29 May 2015.
See Usage Policy.
29Kb

- 10-20 requests/month
- ~8900 theses, ~1.4M downloads in 2014

Douglas, Taking the Plunge: Requiring the ETD, ETD 2003, Berlin, Germany. http://resolver.caltech.edu/CaltechLIB:2003.003
Coles & Johnson (2010) *Moving Electronic Theses from ETD-db to EPrints: The Best of Both Worlds*
http://resolver.caltech.edu/CaltechLIB:2010.001

Download ratio of 157:1. Alumni who graduated before the ETD requirement occasionally request to have their dissertations scanned and made available, but the majority of requests are from scholars from around the world. When a request comes in, library staff attempt to contact the author and have them fill out a permissions form, after which the dissertation is scanned and made freely available. People want to read our work, as evidenced by the requests.

CaltechTHESIS Outreach



http://libguides.caltech.edu/theses

Caltech Library

Kathy Johnson, our Institutional Repository Librarian, teaches several classes each term on the overall process for thesis depositing and related graduation requirements. She also maintains an online LibGuide to assist students and answer questions. (Nice to see the recognition by Tony of ETDs.)

So, What *is* Value?

Downloads? Citations?
Service to the community? Cost justifications?

http://www.caltech.edu/content/mission-statement
"The mission of the California Institute of Technology is to **expand human knowledge** and benefit society through research integrated with education. We investigate the most challenging, fundamental problems in science and technology in a singularly collegial, interdisciplinary atmosphere, while educating outstanding students to become creative members of society."

http://www.caltech.edu/content/our-vision
"Caltech was founded on the premise of **learning through discovery**, and cultivates an environment where scientists and engineers can pursue solutions to the world's greatest challenges."

Caltech Library

So we've talked about what CODA is, and what it provides, and why. So the next questions is, "so what?" What is value / impact? Like Robert mentioned, are these easy-to-measure metrics useful beyond justifying a political agenda? And why does it matter so much in academia? Everything needs to have its cost justified, from subscriptions to salaries. The cost for CODA is in staff time, so is it worth it to provide this resource for the world? Reminiscent of the CCDC, influencing future research, here are Caltech's Mission and Vision statements. Advancing knowledge for it own sake is what academia is really about, not treating faculty as factories. The demand for our materials demonstrates their value – things won't get used if they aren't out there. And like the desktop CSD, there may be applications we may never know about. These tools also serve faculty needs as well in terms of promotional material and citation list management, so internal value should not be overlooked.

Future Work

Refine tools

Keyword searching, heatmaps, etc.

Citation rate vs. AUTHORS posting date

Other measures of value and impact?

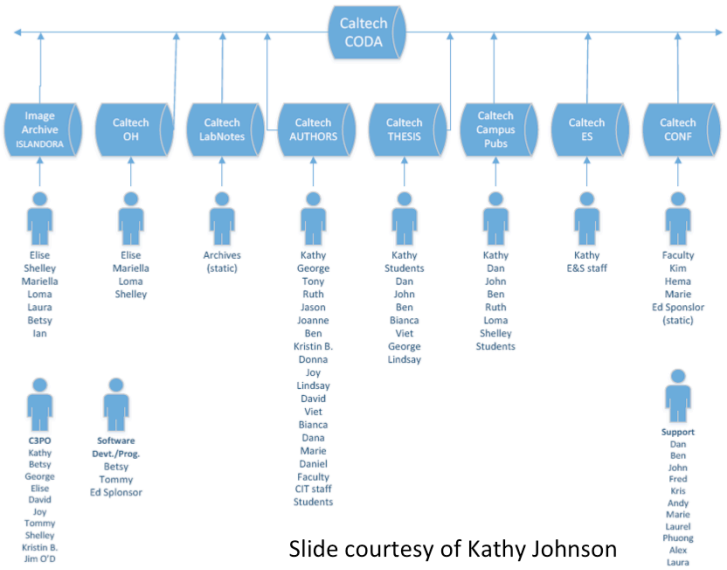
Are we really talking about engagement?

Acknowledgements

Betsy Coles
Kathy Johnson
Caltech Library

Future work – repository software still may have a long way to go to get granular access data.
Engagement – Tony's example of someone taking his work further and really engaging with it (versus something that could be considered "passive" like citations.)
A huge debt of thanks goes to Betsy Coles for wrangling the data out of our Eprints server, which allowed me to generate the heat maps. Kathy Johnson, along with my co-author George, teach several classes each term on CODA. And finally the Library for support. And of course…

Acknowledgements



Slide courtesy of Kathy Johnson

Caltech Library

249th American Chemical Society National Meeting, Denver CO
March 22, 2015

21

…everyone who works so hard on CODA.